A blockchain based federated learning for message dissemination in vehicular networks

Article (Accepted Version)

A Blockchain based Federated Learning for Message Dissemination in Vehicular Networks

Ferheen Ayaz, *Graduate Student Member, IEEE,* Zhengguo Sheng, *Senior Member, IEEE,* Daxin Tian, *Senior Member, IEEE* and Yong Liang Guan, *Senior Member, IEEE*

***Abstract*—Message exchange among vehicles plays an important role in ensuring road safety. Emergency message dissemination is usually carried out by broadcasting. However, high vehicle density and mobility lead to challenges in message dissemination such as broadcasting storm and low probability of packet reception. This paper proposes a federated learning based blockchain-assisted message dissemination solution. Similar to the incentive-based Proof-of-Work consensus in blockchain, vehicles compete to become a relay node (miner) by processing the proposed Proof-of-Federated-Learning (PoFL) consensus which is embedded in the smart contract of blockchain. Both theoretical and practical analysis of the proposed solution are provided. Specifically, the proposed blockchain based federated learning results in more vehicles uploading their models in a given time, which can potentially lead to a more accurate model in less time as compared to the same solution without using blockchain. It also outperforms other blockchain approaches in reducing 65.2% of time delay in consensus, improving at least 8.2% message delivery rate and preserving privacy of neighbor vehicle more efficiently. The economic model to incentivize vehicles participating in federated learning and message dissemination is further analyzed using Stackelberg game. The analysis of asymptotic complexity proves PoFL as the most scalable solution compared to other consensus algorithms in vehicular networks.**

***Index Terms*—blockchain, federated learning, smart contract.**

I. Introduction

TRADITIONAL Vehicular Ad-hoc NETwork (VANET) is growing into Internet-of-Vehicles (IoV) to manage large amounts of data transmission, computation & storage and to meet the increasing requirements of infotainment and road safety [1]. An IoV enables Vehicle-to-Everything (V2X) communications including Vehicle-to-Vehicle (V2V) and Vehicle-to-Infrastructure (V2I) communications. V2I communication usually refers to an infrastructure dependent VANET, where a cellular base station or a Road Side Unit (RSU) is used to provide a real-time and reliable traffic information. However, a large number of RSUs to provide full coverage in urban areas and high traffic densities require huge installation and maintenance cost [2]. Therefore, effective and reliable infrastructure-less V2V communication is necessary in emergency situations such as accidents and traffic jams, so that traffic information can be exchanged in real time, even if RSU is out of reach. In V2V communications, multi-hop relaying is one of the challenges to successfully deliver a message over a wide area. Optimal relay selection mechanisms result in better coverage, more reliable connectivity and less communication overhead [3]. Various intelligent relay selection schemes depending on a vehicle's distance from predecessor, moving direction, speed and propagation loss in environment have been proposed using fuzzy logic [4] or machine learning algorithms [5]. Existing literature shows improved packet delivery ratio by machine learning algorithms in multi-hop V2V communications [6]. However, artificial intelligence methods require huge processing power and are often not suitable for a fully distributed architecture [7].

In a traditional centralized architecture of machine learning, the data collected by mobile devices is uploaded and processed in a cloud based server to produce inference models [8]. With potentially large number of autonomous vehicles, where real-time decisions have to be made within a restricted time period, a cloud-centric approach is unable to offer acceptable latency and scalability. Also, a centralized architecture requires full connectivity which is challenging for vehicular networks. Federated learning (FL) is a distributed machine learning approach, in which mobile devices collect data and train their individual machine learning or deep learning models, called local models. They send their local models (i.e., models' weights) to an aggregator. The aggregator averages local models and produces a global model. Mobile devices further train the global model individually to create updated local models and submit them to aggregator. The steps are repeated in multiple iterations until a desired accuracy of global model is achieved [9]. FL is considered as a feasible solution for safety and time critical applications involving autonomous vehicles [10].

Despite offering a distributed approach, FL still relies on a central aggregator. Furthermore, it needs a sustainable economic model to incentivize mobile devices based on their contributions and prevent adversary attacks. For example, in IoV, a malicious vehicle may deliberately modify data, causing poisoning attack [11] or a selfish vehicle may not cooperate

This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 101006411. This research was also supported in part by the Newton Advanced Fellowship under Grant No. 62061130221, National Natural Science Foundation of China under Grant No. U20A20155, 61822101 and 62173012, in part by the Beijing Municipal Natural Science Foundation under Grant No. L191001 and A*STAR under its RIE2020 Advanced Manufacturing and Engineering (AME) Industry Alignment Fund–Pre Positioning (IAF-PP) (Grant No. A19D6a0053).

F. Ayaz and Z. Sheng are with the Department of Engineering and Design, University of Sussex, Brighton, BN1 9RH, U.K. e-mail: (f.ayaz@sussex.ac.uk, z.sheng@sussex.ac.uk).

D. Tian is with School of Transportation Science and Engineering, Beihang University, Beijing, 100191, China. e-mail: (dtian@buaa.edu.cn)

Y. L. Guan is with the School of Electrical and Electronic Engineering, Nanyang Technological University, 639798, Singapore. e-mail: (eylguan@ntu.edu.sg)

TABLE I: FL issues & blockchain solutions.

FL issue in vehicular networks	Blockchain-based solution
Requires central aggregator	Independent of third party
Lacks economic modeling	Manages cryptocurrency based incentives
Requires adversary control	Provides security by smart contract

in data collection resulting in inaccurate weights of a local model. Blockchain can be used with FL to provide a decentralized solution, for managing incentives and ensuring security and privacy in a trustworthy manner [12]. A blockchain is a distributed ledger of immutable blocks which are added after undergoing a set of rules called consensus [13]. Due to its decentralized nature, blockchain complements both FL and IoV [14]. Furthermore, smart contracts, which are self executing scripts stored in blockchain to enforce a set of rules, allow automation of multi-step processes and interaction among mobile devices [15]. Therefore, they can be used to set rules for protecting FL from adversary and security attacks. The process of transaction verification in blockchain can also be utilized to validate local models in FL [16]. Table I summarizes the current issues of FL in IoV and corresponding solutions provided by blockchain.

Practical implementation of blockchain in vehicular networks is challenging. Due to limited connectivity duration in V2V communications, moving vehicles may not always have an updated blockchain ledger, which leads to possibility of multiple blocks added in parallel, called forks, as shown in Fig. 1 (a). With presence of forks, it is difficult for all vehicles in a network to attain a synchronized linear structure of ledger. As a common practice, blockchain picks one of the parallel blocks to continue, and meanwhile, disqualifies other forking blocks by longest chain acceptance protocol [17]. Forks also lead to creation of malicious chains [18]. To address this issue, the hierarchical structure of blockchain is proposed for vehicles in [19]-[20]. In a hierarchical structure, there are two types of blocks: keyblock and microblock. Instead of a linear ledger, microblocks representing off-chain transactions are added in parallel, whereas keyblocks are main blocks which are appended horizontally in a blockchain by a leader or a central node, for example, RSU. As shown in Fig. 1 (b), parallel addition of microblocks does not disturb the main linear ledger and forks are not disqualified but accepted as off-chain micro-transactions recorded in a decentralized manner.

In this paper, we propose a decentralized message dissemination solution using a hierarchical blockchain based FL process. The vehicles train local models and RSU acts as aggregator to consolidate the global model. The process uses blockchain for updating local models in a decentralized manner. The main contributions of the paper are:

- We propose a blockchain based FL process in vehicular networks, where a smart contract is used to control adversary attacks by malicious and selfish vehicles. Lower Mean Squared Error (MSE) in less number of iterations is achieved by the global model produced through FL if security check is enforced by the smart contract. An economic model to incentivize relay nodes and vehicles participating in FL process is also presented together with its analysis using Stackelberg game.
- Theoretical and simulation analysis in presence and absence of blockchain are presented. In a given time slot, the number of local models uploaded through blockchain based FL are higher than a centralized approach without blockchain, which concludes that the proposed solution can achieve greater accuracy within less time.
- We propose a Proof-of-FL (PoFL) consensus for a blockchain-based multi-hop relay selection scheme in V2V communications. PoFL results averagely in a reduced time delay per hop by 65.2%, an improved message delivery ratio by at least 8.2% and a more privacy-preserving approach as compared to other blockchain approaches suitable for message dissemination.

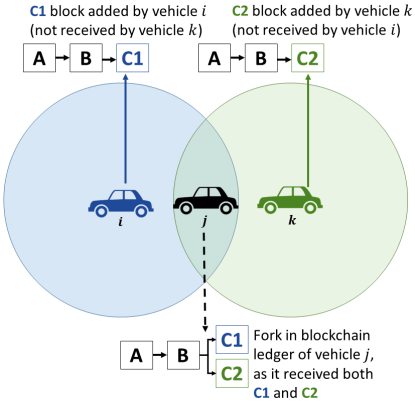
The rest of the paper is organized as follows. Section II describes related work. Section III explains the proposed solution of blockchain based FL and PoFL based message dissemination with discussion on its privacy feature. Section IV theoretically analyzes training capacity of FL and the proposed economic model. Simulation results and conclusion are presented in Section V and Section VI respectively.

II. Related Work

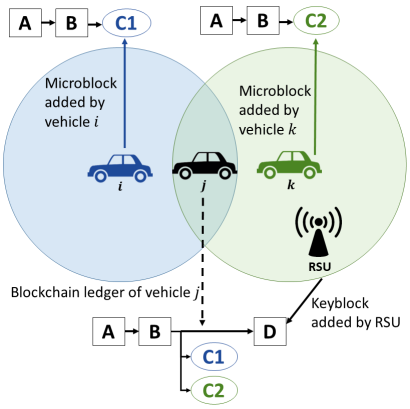
A. Multi-Hop Relay Selection

Relay selection in V2V communications plays a crucial role in message broadcasting. An inappropriate relay may cause unacceptable latency or sometimes failure in delivering a message to a desired number of vehicles or area. Probabilistic calculations are usually used to predict either the distance [21] or link stability [22] of a relay node. In [20], a Proof-of-Quality-Factor (PoQF) is established using probabilistic estimation of both distance and Signal to Noise Interference Ratio (SINR) as a merit of relay node selection. However, probabilistic predictions rely on certain approximations, for example, number of vehicles within a transmission range, which may not be highly accurate with varying speeds. In [23] and [24], the combination of distance of a vehicle from previous sender and channel quality parameters are used to determine link stability for relay node selection. It is crucial to set weights of all parameters according to their impacts on message delivery in a network.

To make relay selection more adaptive to network changes, artificial intelligence based mechanisms are designed. Fuzzy logic has been used in [4] and [25], which makes decision according to distance, moving direction and speed of vehicles. However, fuzzy logic is also dependent on thresholds and weights to be set in the rule base for making inferences. In [5], satellite images are used to detect buildings and obstacles to enable machine learning driven channel characterization. The path with lowest propagation loss is used for message dissemination in [5]. RSU assisted deep learning based technique is developed for relay selection in [7]. It is pointed out that machine learning and deep learning algorithms require large processing power to handle huge amount of data and therefore they must require V2I communications and infrastructure support for implementation.



(a) Fork (parallel blocks) in linear blockchain.



(b) Parallel microblocks and linear keyblock.

Fig. 1: Parallel addition of microblocks to resolve forks in blockchain in vehicular networks.

TABLE II: Multi-hop relay selection challenges and solutions offered by blockchain-based FL.

Approach	Challenge	Solution
Probabilistic prediction [20], [22]	Assumptions / rules are not adaptable to network changes	Local models trained with different networks and the global model can cater network changes
Fuzzy logic [4], [25]		
Machine learning [5], [7]	Huge data have to managed centrally	Distributed learning and decentralized storage
Any scheme without incentives [21]	Relay nodes may act selfish	Blockchain incentives for motivation

B. FL in Vehicular Networks

FL is suggested as a promising technique to securely train intelligent models across smart cars [10] and Unmanned Aerial Vehicles (UAVs) [26]. It has the feature of reducing network latency by dividing training task among network edges. In cellular-V2X (C-V2X) communications, FL is proposed to reduce failure probability by intelligently offloading high computation tasks to nearby base stations [27]. Resource allocation and sharing in C-V2X by FL among vehicles has promised better coverage and Quality-of-Service (QoS) in [28]. FL and fog-assisted V2X is presented in [29] to improve driving experience of autonomous vehicles by providing user-end services, for example, car sharing, intelligent parking allocation, infotainment and e-commerce applications. In [30], FL is used to tackle energy transfer issues of electric vehicles at charging stations and has resulted in improved accuracy of energy demand prediction. FL assisted blockchain is proposed in [31] to adjust block arrival rate in order to reduce communication latency and consensus delay among vehicles. Applications of FL in vehicular networks are summarized in [32] and most of the recent applications focus on resource management, performance optimization in computing tasks and user-end services. However, FL can also be promising in message delivery and relay node selection. Table II summarizes the challenges of existing multi-hop relay selection schemes and solutions offered by blockchain-based FL.

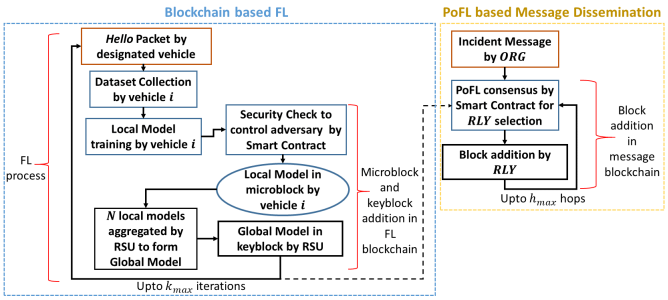
C. Economic Modeling in FL

Economic models to strategize incentives and to promote mobile devices for producing reliable local models have been developed. For a blockchain based FL in [31], the authors have suggested to incentivize vehicles for both model training and block mining. A joint price and reputation based economic model is proposed in [11] to incentivize devices according to the size of data contributed and prevent poisoning attack. The economic model is analyzed using Contract Theory. In Contract Theory, the contracts are formed between a payer and a service provider (i.e., devices training local models) before initiation of FL process. FL among vehicles for image classification is proposed in [33] and contracts are formulated to incentivize vehicles in proportion to the number of images used and amount of computation resources consumed.

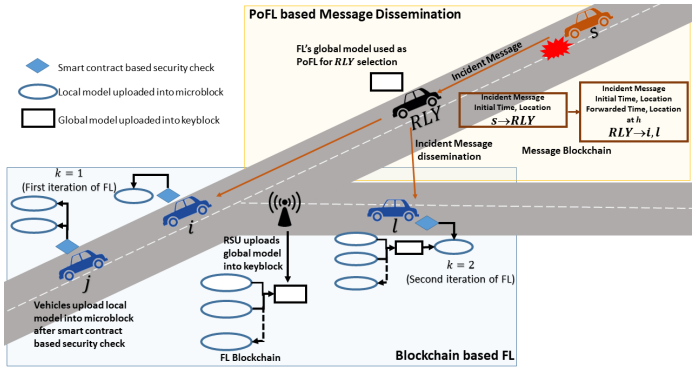
Stackelberg game approach is used in [34] and [35] to analyze the actions of players when incentives are distributed after FL iterations are completed. If relay nodes are involved in incentive distribution among vehicles, the economic model is more suited to be analyzed using Stackelberg game model. Due to varying speed and position of vehicles, it is practically better to select an appropriate relay node after a message is originated. Therefore, analysis using Stackelberg game model is a more feasible option for multi-hop relay selection scheme than Contract Theory, because formation of contract prior to FL process initiation or relay selection cannot be materialized. The existing literature assumes information asymmetry, i.e., the payer is not aware of the amount of contributions (for example, data size) upon which the payment is to be made. However, with public blockchain, where stored transactions are visible to every member of blockchain, FL can be a case of symmetric information, i.e., the relevant information is known to all associated members.

III. The Proposed Solution

As illustrated in Fig. 2, the proposed solution consists of a blockchain-based FL process and a solution for multi-hop relay selection. The FL process is aimed to form a global model which is later used as a consensus to select a relay node (RLY) when an incident message is originated by a vehicle.



(a) The proposed steps for vehicles and RSU.



(b) Blockchain-based FL and Message Dissemination.

Fig. 2: The proposed solution of blockchain based FL for message dissemination in vehicular networks.

TABLE III: Key notations.

Notation	Definition
ORG	Originator vehicle
RLY	Relay node vehicle
k	Iteration index
$\boldsymbol{w}_x^k$	Weights of model x (local or global) at k^{th} iteration
$L(\boldsymbol{w}_x^k)$	Loss function of model x at k^{th} iteration
TS	Time slot to upload local models
R	Transmission range
λ_{MB}	Microblock arrival rate per second
λ_V	Vehicle density per m^2
μ_d	Average distance of vehicles from RSU
μ_v	Average speed of vehicles
$E(.)$	Expected value
N	No. of vehicles participating in FL
N_B	No. of vehicles uploading local models via FL blockchain
N_{WB}	No. of vehicles uploading local models without blockchain
N_V	No. of vehicles with RSU in transmission range
N_{MV}	No. of moving vehicles reaching RSU
N_{RLY}	No. of relay nodes
s_i	Data size of vehicle i
I	Incentive
$C(s_i)$	Cost of training on data of size s_i
U_x	Utility of x (vehicle i or RLY)
α_i	Cost coefficient of vehicle i
β	Compensation
p_m	Probability of using s_m
M	Number of possible s_m
Dataset	
$d_{i,x}$	Distance between vehicle i and x (vehicle or RSU)
$dir_{i,s}$	Direction of vehicle i w.r.t sender s
v_i	Speed of vehicle i
h	Hop index
γ_i	Traffic density in transmission range of vehicle i
N_A	No. of acknowledgment messages

Overall, the proposed approach consists of two major parts:
(1) FL integrated with blockchain, where vehicles take part in a blockchain based FL process to form a global model for relay node selection, and
(2) PoFL based message dissemination, where the global model produced in first part is used to find vehicles' eligibility to become a relay node. Table III lists the key notations used in this paper.

A. Blockchain based FL

a) Elements asscociated with FL:

- ***Hello* Packet by designated vehicle:** A Central Authority appoints some designated vehicles to regularly originate a *Hello* packet and share their position to initiate dataset collection by vehicles participating in FL. Only the designated vehicles are allowed to originate *Hello* packets. The motivation behind designated vehicles is two fold: first is because they are trusted by Central Authority to honestly send their actual position without any malicious change and second is because the encrypted identities of designated vehicles are already shared with other vehicles, so *Hello* packet from any other identity is not recognized by the network. Designated vehicles can either be representatives of Central Authority or selected from the existing network based on their trust ratings. Calculations and storage of trust ratings are out of the scope of this paper but can be managed by a separate blockchain.
- **Dataset:** It refers to the data samples collected by vehicle i for training local model. In the proposed solution, dataset collected by vehicle i includes multi-hop relay selection parameters. After forwarding a *Hello* packet, dataset collected by vehicle i consists of the following parameters mentioned in Table III: $d_{i,s}$, distance from sender s (designated vehicle or previous relay node). $dir_{i,s}$, moving direction (either towards or away from sender s). v_i, speed at the time of forwarding message. h, hop number. γ_i, traffic density within its transmission range and N_A, number of acknowledgments received as the *score* of relaying. γ_i in dataset can be pre-specified by Central Authority or estimated by counting average

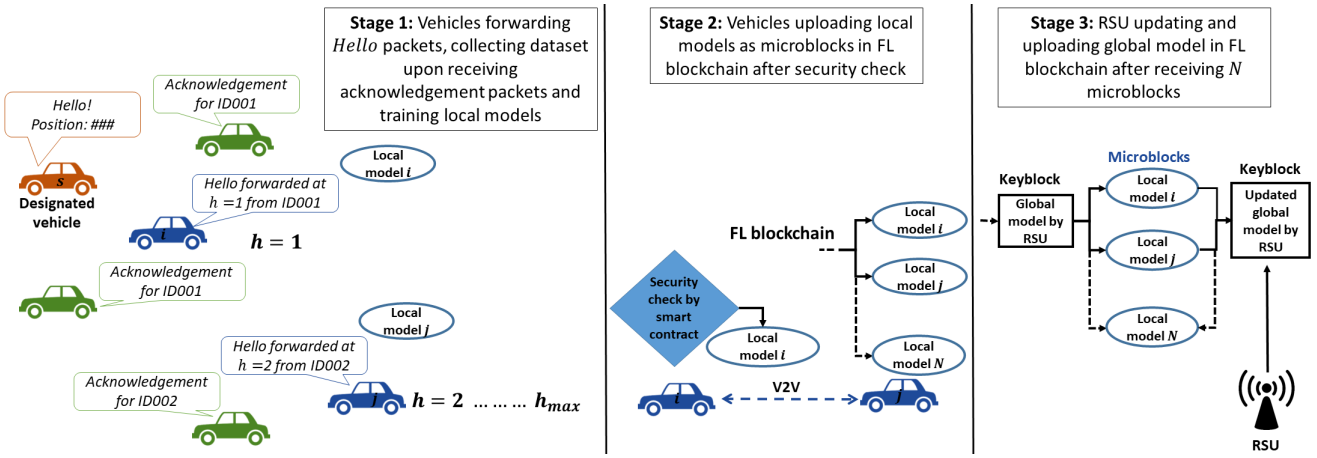


Fig. 3: The proposed stages in blockchain based FL.

number of vehicles sending beacon messages per meter [36]. The process of dataset collection is explained in detail later in this section. The motivation behind individual dataset collection by each vehicle is two-fold. First, the datasets are not shared among vehicles to protect privacy. Second, vehicles training their local models on different dataset brings diversity in learning, so an efficient global model is created .

- **Local Model:** Each vehicle i participating in FL trains a Deep Neural Network based local model.
- **Global Model:** It is an aggregated model of N local models, where N also refers to the number of vehicles participating in FL. In our proposed system, RSU collects N local models and aggregates them to form a global model.

b) Adversary: We consider the following adversary threats:

- **Malicious Vehicles:** They may deliberately change or inject false data so that local model is not trained accurately. This phenomenon is also known as poisoning attack [11].
- **Selfish Vehicles:** They may not send acknowledgment messages despite receiving forwarded messages. Therefore, N_A cannot be recorded correctly during dataset collection, leading to an inaccurate local model produced by vehicle i.

c) FL Blockchain and its Components: FL blockchain is a blockchain used by vehicles and RSUs to store local and global models as blocks. Its main components include

- **Security Check:** It is a machine learning algorithm embedded in smart contract of FL blockchain to detect adversary before a local model is uploaded as a block by vehicle i.
- **Microblock:** A local model is stored in FL blockchain as a microblock after undergoing a security check. A microblock is added in parallel to other microblocks, all containing hash of previous keyblock.
- **Keyblock:** A global model is stored in FL blockchain at RSU in the form of a keyblock, containing hashes of previous N microblocks.

Algorithm 1 FL Algorithm for vehicle i

Input: *Hello* Packet, N vehicles
Output: Global Model

while $h \leq h_{max}$ **do**
  Generate random waiting time
  while Time elapsed $\leq$ random waiting time **do**
    if Forwarded *Hello* packet received at h **then**
      break
    end if
  end while
  if Forwarded *Hello* packet not received at h **then**
    Forward *Hello* packet
    Count acknowledgment packages into N_A
    Record $d_{i,s}$, v_i, $dir_{i,s}$, γ_i, N_A in dataset
    break
  else
    $h = h + 1$
  end if
end while
if data size $==$ s_i **then**
  Train local model
else
  Go to 1
end if
while $k \leq k_{max}$ **do**
  Upload local model through smart contract
  Receive updated global model
  Re-train local model
  $k = k + 1$
end while

As shown in Fig. 3, the proposed blockchain based FL consists of the following three stages:

1) Stage 1: Dataset Collection and Local Model Training: In this stage, vehicles collect dataset for training. Upon receiving a *Hello* packet from a designated vehicle, a vehicle i which aims to collect dataset, generates a random waiting time. When the waiting time is complete, it forwards a *Hello* packet with its encrypted identity. The reason behind a random waiting time

is to prevent multiple vehicles from transmitting at the same time and avoid packet collision. The limits and probability distribution of random waiting time are described in [20]. The vehicles which receive the forwarded *Hello* packet for the first time share their acknowledgment. An acknowledgment packet contains encrypted identity of vehicle i, so that it can collect dataset. A vehicle j, which participates in FL, will broadcast the received *Hello* packet again after a random waiting time. This process continues up to a specified number of hops, h_{max}, as shown in Algorithm 1. Each vehicle produces a local model based on Deep Neural Network with 7 hidden layers and 256 neurons in each layer.

2) Stage 2: Security Check and FL Blockchain Update: A vehicle i shares its local model with the network by adding it into FL blockchain as a microblock. It is added after passing through a security check performed by the smart contract embedded in FL blockchain. The proposed security check employs a machine learning algorithm called Isolation Forest [37] to detect anomaly in a local model caused by adversary. Isolation Forest is used because it only requires a small number of samples for training. A true sample of dataset is provided by the Central Authority for its initial training. Later, it can be used in a fully unsupervised manner to detect anomaly. Moreover, it is computationally efficient and has low memory requirement [38]. We have used the security check in three ways. Firstly, the security check on dataset is conducted by finding anomalies in dataset of each vehicle. Secondly, the security check performs anomaly detection on weights of local models. If a malicious vehicle i deliberately changes its dataset for training its local model but shares a true dataset in smart contract, the adversary attack will be detected by anomaly detection on weights. Thirdly, the security check on both dataset and weights is performed. If local models successfully pass the security check, they are added in FL blockchain in the form of parallel microblocks. The microblock announcement is broadcasted by vehicle i and the receiving vehicles will then update their copy of FL blockchain. Vehicles can exchange new microblock updates with their neighbors regularly.

3) Stage 3: Global Model Aggregation: Whenever a vehicle i finds an RSU available in its transmission range, it shares its updated copy of FL blockchain. When N microblocks are received by RSU in FL blockchain, it aggregates local models into a global model and uploads it into a keyblock.

All stages are repeated at each iteration. The goal is to repeat the process up to k_{max} iterations for minimizing global loss function $L(\boldsymbol{w}_G^k)$, which is defined as

$$L(\boldsymbol{w}_G^k) = \frac{1}{N}\sum_{i=1}^{N} L(\boldsymbol{w}_i^k). \tag{1}$$

where $\boldsymbol{w}_G^k$ are weights of global model, $L(\boldsymbol{w}_i^k)$ is the loss function of local model i and $\boldsymbol{w}_i^k$ are its corresponding weights at k^{th} iteration. Neural networks commonly use MSE as the loss function [10]. The value of k_{max} is adjusted by Central Authority to achieve the minimum possible or desired $L(\boldsymbol{w}_G^{k_{max}})$ [9].

Algorithm 2 Message Dissemination Algorithm for vehicle i

Input: Incident message, global model
Output: New block announcement in message blockchain

while $h \leq h_{max}$ **do**
  Compute *score* from global model
  timer expiry limit = $1/score$
  while Time elapsed $\leq$ timer expiry limit **do**
    if New block announced **then**
      break
    end if
  end while
  if New block not announced **then**
    Announce block
    break
  else
    $h = h + 1$
  end if
end while

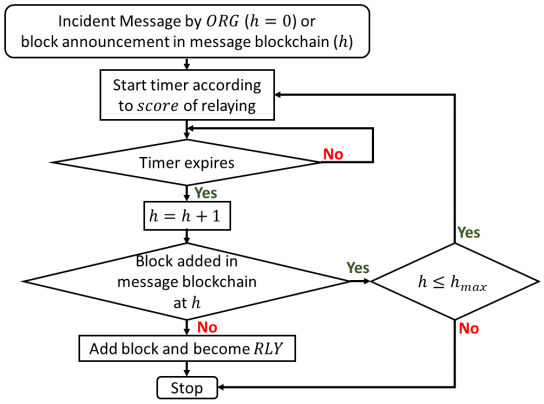


Fig. 4: Flowchart of actions by vehicle i according to PoFL based Message Dissemination.

B. PoFL based Message Dissemination

The main elements of this part include

- **Incident Message:** It is a message initiated by an originating vehicle (ORG) in an emergency situation, for example, incident or traffic jam. It contains time and position of incident and encrypted identity of ORG.
- **PoFL:** It is the consensus to select RLY for forwarding an incident message. Global model contained in the latest keyblock of FL blockchain is used as PoFL. It is run by a smart contract of message blockchain.
- **Message Blockchain:** It contains history of incident messages. The selected RLY adds a block in the message blockchain containing the forwarded incident message. The block also contains time, location and encrypted identity of the RLY which adds the block. The motivation behind message blockchain is to record forwarded incident messages as immutable blocks and avoid discrepancies in allocating incentive to RLY at each hop.

When an incident message is initiated by ORG, all receiving vehicles attempt to become the RLY by competing through PoFL consensus. Each vehicle i runs PoFL consensus to find

TABLE IV: Parameter-sharing required from neighbor vehicles in multi-hop relay selection.

Approach	Position	Speed	Other Parameters
Deep learning [7]	✓	✓	Transmission power
Fuzzy logic [25]	✓	×	×
Probabilistic prediction [22]	✓	✓	×
Link Stability [23]	✓	×	Channel quality
PoQF [20]	✓	✓	×
PoFL	×	×	×

its *score* of being RLY, as shown in Algorithm 2. PoFL is aimed to assign the highest *score* to the most appropriate vehicle as RLY. The smart contract starts a timer whose length is inversely proportional to the *score* of vehicle i. As shown in Fig. 4, a block is added in the message blockchain and a block announcement with the forwarded incident message is initiated by vehicle i if its timer first expires. In this case, vehicle i is assigned as a relay node RLY at $h = 1$. All other vehicles continue to compete for becoming RLY at further hops until the message is forwarded up to h_{max} number of hops.

a) Privacy of PoFL based Message Dissemination: Table IV lists the parameters required to be shared by neighbor vehicles in various multi-hop relay selection approaches. The position, speed and heading direction of vehicles are regularly shared in VANETs using beacon messages and thus create a threat to privacy [39]. The proposed approach does not require such information from all neighbor vehicles and can therefore be considered as a privacy-preserving solution. The position and direction of only sender is required for dataset collection in blockchain based FL and for calculating *score* of relaying in PoFL based message dissemination. However, identities of vehicles are kept anonymous using encryption. Disclosure of identities through brute force attack is a possibility but it will not be very effective for the attacker. Due to high time complexity of brute force attack [40], the position, speed and direction of a vehicle will be changed until its identity is disclosed. In case of high probability of brute force attack, a private blockchain with only trusted vehicles [23] or timely refreshing of cryptographic identities is recommended [41].

IV. Theoretical Analysis

A. Training Capacity of FL

This subsection is aimed to analyze the capacity of FL blockchain to complete one FL iteration in a given amount of time, compared with the same process without blockchain. FL without blockchain is referred to as a centralized approach in which each vehicle i submits its local model directly to RSU instead of uploading it into FL blockchain. As the convergence performance of FL improves with increasing number of local models [42], FL via blockchain is expected to achieve greater accuracy, provided the number of uploaded local models are higher as compared to the process carried out without blockchain within the same time period.

1) FL with blockchain: Let TS be a time slot in which a vehicle i is required to upload its local model as microblock

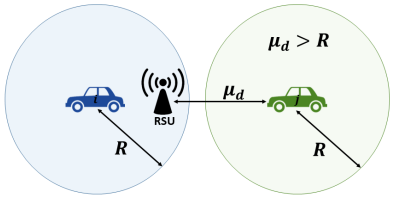


Fig. 5: Distance between vehicle and RSU.

in FL blockchain, after it has completed training and passed its local model through security check. Let λ_{MB} be the microblock arrival rate at RSU or throughput in microblocks/s. Detailed derivation of λ_{MB} can be found in [20]. If microblock arrival is modeled using Poisson distribution as defined in [18], the expected number of vehicles able to upload thier local models via FL blockchain in TS can be given as [43]

$$E(N_B) = \sum_{l=1}^{\lambda_{MB}TS} l e^{-\lambda_{MB}TS} \frac{(\lambda_{MB}TS)^l}{l!}. \tag{2}$$

2) FL without blockchain: If vehicles are required to upload their models directly to RSU without blockchain, then it is necessary that either RSU is in their transmission range or they are able to reach towards RSU within TS. Consider a general and dynamic movement of vehicles, the position of vehicles on road follows Poisson distribution, λ_V vehicles/m^2 is assumed as the density of vehicles on a two dimensional road segment with no separation of lanes [44], the expected number of vehicles with RSU in their transmission range is

$$E(N_V) = \sum_{l=1}^{\lambda_V \pi R^2} l e^{-\lambda_V \pi R^2} \frac{(\lambda_V \pi R^2)^l}{l!}, \tag{3}$$

where transmission range is assumed as a uniform circle with radius R. Similarly, the expected number of moving vehicles with RSU not currently in their transmission range but can travel to reach RSU within TS is

$$E(N_{MV}) = \sum_{l=1}^{\frac{TS}{\mu_d - R}\mu_v} l e^{-\frac{TS}{\mu_d - R}\mu_v} \frac{(\frac{TS}{\mu_d - R}\mu_v)^l}{l!}, \tag{4}$$

where $\mu_d > R$ is the mean distance of those vehicles from RSU which do not have RSU within their transmission range, as shown in Fig. 5 and μ_v is their average speed. Therefore, the expected number of vehicles able to upload their local models to RSU without blockchain during TS is

$$E(N_{WB}) = E(N_V) + E(N_{MV}). \tag{5}$$

B. Economic Model

In this subsection, we define an economic model of payment to vehicles contributing in FL and message dissemination. The feasibility of economic model is analyzed by investigating strategic behavior of RLYs and vehicles participating in FL based on their expected utilities through Stackelberg game model.

TABLE V: Reward gained and payment made by players.

Player	Gains	Pays
ORG	None	$N_{RLY}\beta log(1+I)$ to $RLYs$
RLY	$\beta log(1+I)$ from ORG	$\sum_{i=1}^{N} Is_i$ to N vehicles
Vehicle i	$N_{RLY}Is_i$ from $RLYs$	$\alpha_i s_i^2$ to train local model

1) Stackelberg Game Formulation: The Stackelberg game model consists of three types of players: ORG, RLY participating in message dissemination and vehicle i participating in FL. For each incident message initiated by ORG, there are N_{RLY} number of $RLYs$ which forward the incident message and N vehicles in the network which train their local models during blockchain based FL. The proposed economic model is formulated as a two-stage Stackelberg game. First, at stage 1, ORG pays reward to $RLYs$ for forwarding message. At stage 2, $RLYs$ pay reward to N vehicles for participating in FL to form a global model of RLY selection. Since, both FL and message dissemination processes are blockchain-based, the contribution of players is stored as immutable timestamped blocks and cannot be altered through cheating. The transactions of incentives are also processed automatically in the form of blockchain based virtual currency through smart contracts.

As shown in Table V, the reward to N vehicles is paid in proportion to the sizes of dataset they have used in training their local models. Assume that the dataset sizes of N vehicles are $\boldsymbol{s} = \{s_1, s_2, ..., s_N\}$. The utility of each vehicle i participating in FL process is

$$U_i(s_i, I) = N_{RLY} I s_i - C(s_i), \tag{6}$$

where I denotes incentive which is constant for every vehicle i and $C(s_i)$ is the computational cost of training a local model on dataset of size s_i and is modeled as a quadratic function, i.e.,

$$C(s_i) = \alpha_i s_i^2, \tag{7}$$

where $\alpha_i > 0$ denotes cost co-efficient of vehicle i [45]. The utility of each RLY is

$$U_{RLY}(\boldsymbol{s}, I) = \beta log(1+I) - \sum_{i=1}^{N} I \cdot s_i, \tag{8}$$

where $\beta log(1+I)$ is paid by ORG for forwarding the incident message. Here $\beta > 0$ and can be assumed as a compensation amount paid to $RLYs$ present in an area affected by an incident or traffic jam caused by ORG.

2) Stackelberg Game Analysis: We consider the case of information symmetry where every RLY knows data size used by each vehicle i prior to forwarding a message.

Definition 1: Assume that s_i^* is the optimal data size for each vehicle i and I^* is the optimal incentive amount per data size paid by each RLY to vehicle i, then (s_i^*, I^*) is the Nash equilibrium point which satisfies the following conditions

$$U_i(s_i^*, I^*) \geq U_i(s_i, I^*), \tag{9}$$

and

$$U_{RLY}(s_i^*, I^*) \geq U_{RLY}(s_i^*, I). \tag{10}$$

Theorem 1: There exists a Nash equilibrium point for a vehicle i with U_i defined in (6).

Proof: For a fixed I^*, U_i is

$$U_i(s_i, I^*) = N_{RLY} \cdot I^* \cdot s_i - \alpha_i s_i^2. \tag{11}$$

The first-order derivative of (11) is

$$\frac{\partial U_i(s_i, I^*)}{\partial s_i} = N_{RLY} \cdot I^* - 2\alpha_i s_i. \tag{12}$$

The second-order derivative of (11) is

$$\frac{\partial^2 U_i(s_i, I^*)}{\partial s_i^2} = -2\alpha_i. \tag{13}$$

Since $\alpha_i > 0$, the second-order derivative of U_i is negative and $U_i(s_i, I^*)$ is a strictly concave function, which proves the existence of Nash equilibrium. □

Theorem 2: There exists a Nash equilibrium point for RLY with U_{RLY} defined in (8).

Proof: The first-order derivative of (8) is

$$\frac{\partial U_{RLY}(\boldsymbol{s}, I)}{\partial I} = \frac{\beta}{1+I} - \sum_{i=1}^{N} s_i. \tag{14}$$

The second-order derivative of (8) is

$$\frac{\partial^2 U_{RLY}(\boldsymbol{s}, I)}{\partial I^2} = \frac{-\beta}{(1+I)^2}. \tag{15}$$

Since $\beta > 0$ and $(1+I)^2 > 0$, the second-order derivative of U_{RLY} is negative and $U_{RLY}(\boldsymbol{s}, I)$ is a strictly concave function, which proves the existence of Nash equilibrium. □

Based on Theorem 1 and Theorem 2, we can state that the unique Stackelberg Nash equilibrium point of our model exists. The Central Authority is responsible to choose values of I and β such that Nash equilibrium points for all N vehicles and $RLYs$ become their best response strategies (i.e, $U_i > 0$ and $U_{RLY} > 0$) and all players are willing to cooperate in the proposed game.

The proposed economic model assumes information symmetry, i.e., all players have complete information about s_i. If private blockchain is used, complete information may not be visible to every player and the economic model will be information asymmetric. In this case, players may predict information through a machine learning method [46] or using probabilistic assumption [45]. Let $\boldsymbol{s} = \{s_1, s_2, ...s_M\}$ be the sizes of dataset used by vehicles in FL and p_m be the probability that a vehicle i uses s_m. (8) can be modified as

$$U_{RLY}(\boldsymbol{s}, I) = \beta log(1+I) - \sum_{m=1}^{M} p_m N I s_m, \tag{16}$$

TABLE VI: Simulation Parameters.

Parameters	Values	Parameters	Values
Simulation Time	300 s	Protocol	IEEE 802.11p
Size of area	2.5 km ×2.5 km	Encryption	SHA-256
Data rate	6 Mbps	s_i	8000
Mobility model	Krauss	Loss function	MSE
Number of RSUs	1	R	250 m
Number of vehicles	100, 200, 300	h_{max}	6
k_{max}	100, 110	μ_v	50 km/hr

TABLE VII: Loss (MSE) of global model after 100 iterations.

s_i	$N = 100$	$N = 200$	$N = 300$
2000	0.19643	0.18724	0.16541
5000	0.17251	0.17021	0.16313
8000	0.15297	0.15101	0.15085

where M is the total number of possible s_m. Similar to Theorem 2, the existence of Nash Equilibrium point can be proved for U_{RLY} defined in (16).

V. Results and Discussion

In this section, we discuss simulation results of the proposed solution using OMNeT++, Python and SUMO (Simulation of Urban Mobility). An open-source framework VeINS (Vehicles In Network Simulation) is used to integrate SUMO with OMNeT++ [47]. Python is employed for carrying out FL using Tensorflow library of machine learning. Python can be embedded into a C++ program by writing an extension module [48]. Since OMNeT++ is a modular C++ based network simulator, it supports dynamic loading of Python script at run time. The simulation parameters used are listed in Table VI.

Fig. 6 shows the loss (MSE) of global model with respect to iteration index k. We present the results with 50% adversarial vehicles, as it is the highest amount of adversary a blockchain solution can tolerate [20]. For the sake of generality, the adversary consists of equal percentage of malicious and selfish vehicles. In all cases, the loss converges to its lowest possible value until 100 iterations. However, this convergence is achieved in less number of iterations with 300 vehicles as compared to 100 vehicles, which means that the maximum accuracy of a global model can be attained faster with more vehicles participating in FL. Fig. 6 (a) shows the loss when no security check is implemented. The convergence rate is slower without security check and takes more iterations than those with security checks, as shown in Fig. 6 (b) - (d). Table VII shows loss of global model after 100 iterations with respect to number of vehicles participating in FL without any adversary or security check. As shown in Table VII, the loss is inversely proportional to both dataset size and number of vehicles.

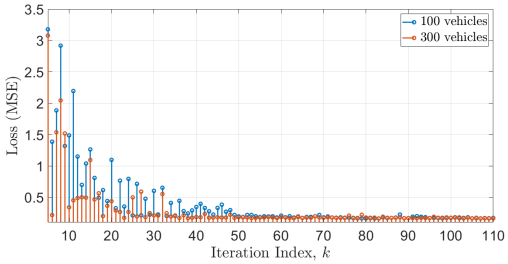
Fig. 7 shows the loss of global model after 100 iterations of FL in presence of adversary. The global loss function is the highest if no security check is employed in smart contract of FL blockchain. Security check on weights results in less loss as compared to security check on dataset in presence of malicious vehicles and oppositely in case of selfish vehicles and combined adversary of equal percentages of malicious and selfish vehicles. Since selfish vehicles only affect N_A in dataset by not sending acknowledgments, such discrepancy is easily detected if security check is applied on dataset only. On the other hand, malicious vehicles can change all parameters in dataset and therefore it is not easy to detect anomaly on such dataset. It shows that security check on weights is more suitable to prevent poisoning attack caused by malicious vehicles and security check on dataset is more appropriate to reduce the effect of selfish behavior. Nevertheless, malicious vehicles may submit a true dataset for security check and upload inaccurate local models using a false dataset. Therefore, in this case, only security check on weights can prevent adversary caused by malicious vehicles. The loss function is minimum with all types of adversary if security check on both dataset and weights is used and is suitable for both malicious and selfish vehicles. As a trade-off, computation time is increased to run security check twice. Fig. 8 shows time consumed per iteration. On an average, security check on dataset or weight takes additional 40 s and security check on both dataset and weights requires 148 s more than an iteration performed without any security check.

TABLE VIII: λ_{MB} and μ_d with respect to λ_V.

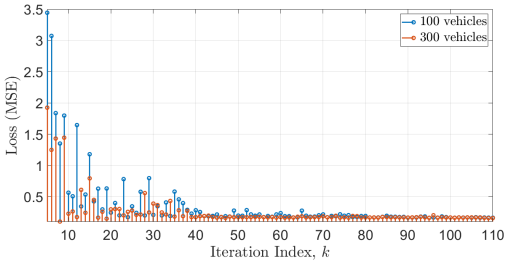
λ_V (vehicle/m^2)	16	32	48
λ_{MB} (microblocks/s)	2.01	1.99	0.98
μ_d (m)	344	298	276

Fig. 9 shows the average number of vehicles over 100 simulation runs which uploaded their local models during TS, with and without blockchain at various λ_V. The simulation results are matched with expected values derived in (2) and (5), confirming our theoretical analysis. λ_{MB} and μ_d change with varying λ_V and are listed in Table VIII. Fig. 9 shows that blockchain based approach results in average 18 more vehicles uploading their local models within same TS compared with the centralized solution in submitting local models directly to RSU without blockchain. This is because a copy of FL blockchain is possessed by each vehicle. A local model by vehicle i can be entered into FL blockchain without depending upon RSU. Subsequently, RSU is able to receive an updated FL blockchain by another vehicle j, containing local models of both vehicle i and vehicle j. Without blockchain, a vehicle i has to travel towards RSU within TS to directly share its local model. In this case, one RSU or small TS may not be sufficient for receiving local models from large number of vehicles. Also, as shown in Table VII, the loss of global model decreases with rising N. It can be concluded that FL blockchain can achieve desired accuracy of a global model faster than FL carried out without blockchain, because FL blockchain enables collection of local models from more vehicles within the same time limit.

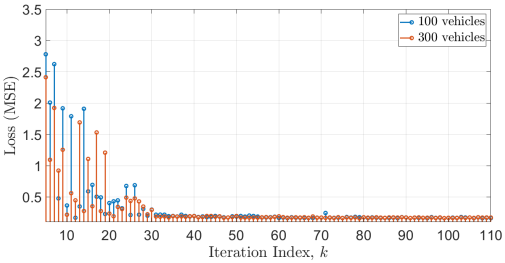
Fig. 10 and Fig. 11 prove Definition 1. The equilibrium points exist with all combinations of I, α_i, β, s_i, N_{RLY} and N. Fig. 10 shows the utility of vehicle i, U_i, participating in a blockchain based FL. As shown in Fig. 10, for a given I^*, there exists only one s_i^* which results in maximum U_i,



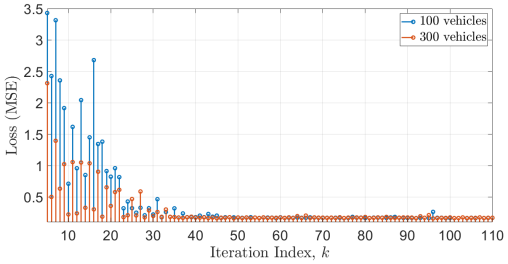
(a) No security Check.



(b) Security Check on dataset.

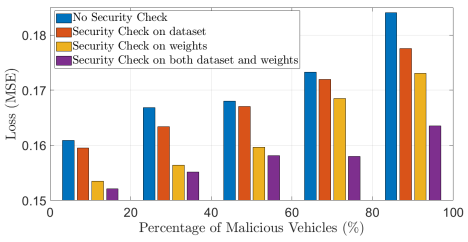


(c) Security check on weights.

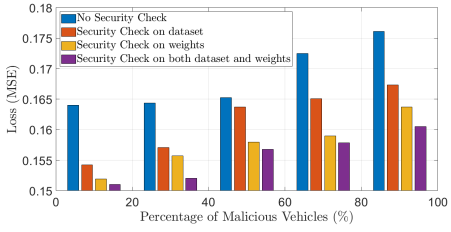


(d) Security check on both dataset and weights.

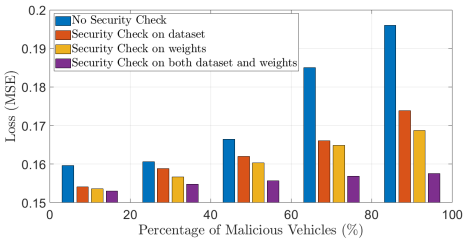
Fig. 6: Loss (MSE) of global model with 50% adversary.



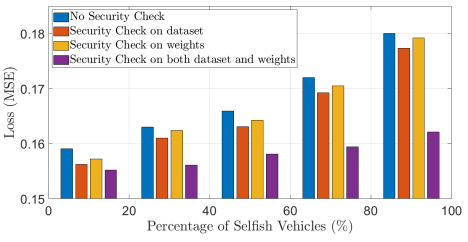
(a) Vehicles: 100, Adversary: Malicious.



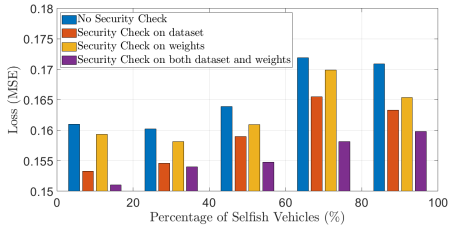
(b) Vehicles: 200, Adversary: Malicious.



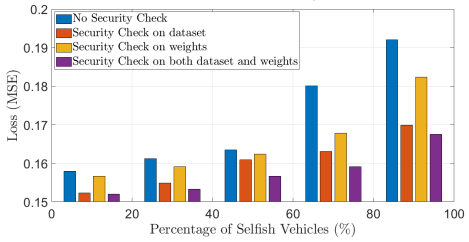
(c) Vehicles: 300, Adversary: Malicious.



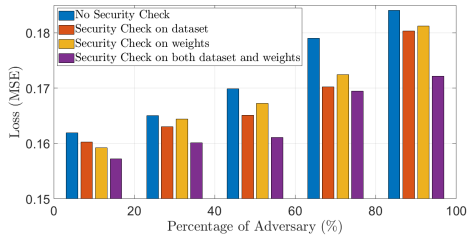
(d) Vehicles: 100, Adversary: Selfish.



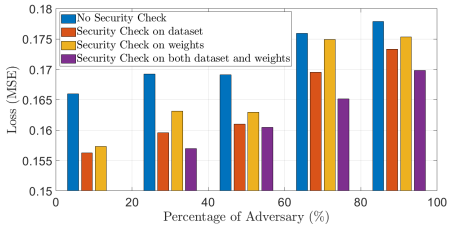
(e) Vehicles: 200, Adversary: Selfish.



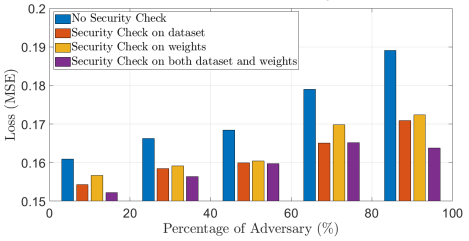
(f) Vehicles: 300, Adversary: Selfish.



(g) Vehicles: 100, Adversary: Equal number of malicious and selfish vehicles.



(h) Vehicles: 200, Adversary: Equal number of malicious and selfish vehicles.



(i) Vehicles: 300, Adversary: Equal number of malicious and selfish vehicles.

Fig. 7: Loss (MSE) of global model after 100 iterations.

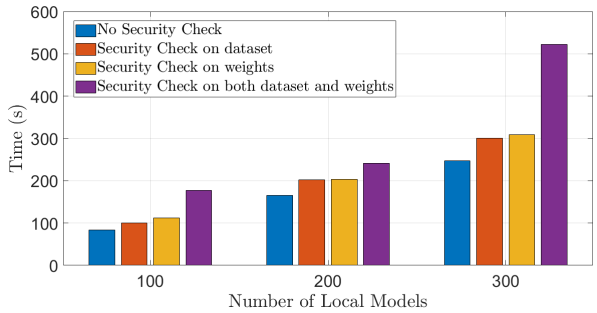


Fig. 8: Average time per iteration.

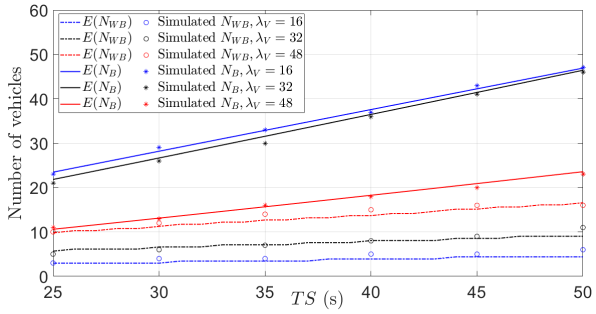


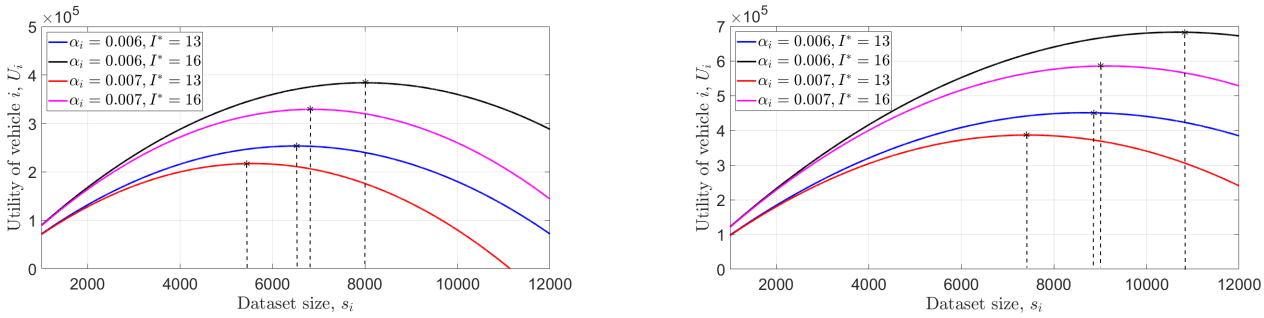
Fig. 9: Number of vehicles uploading local model in TS.

irrespective of N_{RLY}. Similarly, Fig. 11 also shows existence of an equilibrium point with every combination of I, s_i, N and β. A Central Authority can select the value of I^*, which gives both maximum U_i and U_{RLY}. As shown in Fig. 11 (d) and (f), $U_{RLY} <= 0$ for certain values of I, which will motivate $RLYs$ to become selfish. An appropriate value of β can be selected to make $U_{RLY} > 0$ for every N and I. A machine learning model can be used to predict dependent parameters, such as s_i used by each vehicle i, when it is trained using historical information [46]. Thus, the expected utility can be estimated using (8) and the optimum combination of β and I, which results in the best response strategy of $RLYs$. This model can be embedded into smart contract of message blockchain to automate reward distribution independently without Central Authority.

Fig. 12 shows message delivery ratio at varying number of vehicles and percentages in malicious vehicles as a result of 100 simulation runs. Results are also compared with PoQF [20] and another voting blockchain based relay selection method in which an appropriate relay is elected on the basis of its distance from the sender and channel quality parameters [23]. Since both PoQF and voting based approach can tolerate up to half as malicious vehicles in the network and do not consider selfish adversary, the results are presented with up to 50% as malicious vehicles only. Message delivery ratio decreases with number of vehicles in all approaches. Voting based relay selection always results in least message delivery ratio. Results of PoQF are comparable to PoFL with no security check. PoFL with security check on both datasets and weights result in the highest message delivery ratio and outperforms voting based relay selection and PoQF by an average of 25% and 8.2% increase in message delivery ratio respectively. Fig. 12 shows that the improved performance is due to blockchain-based smart contract which executes security check. It proves that a blockchain support is essential for federated learning based message dissemination in presence of malicious vehicles.

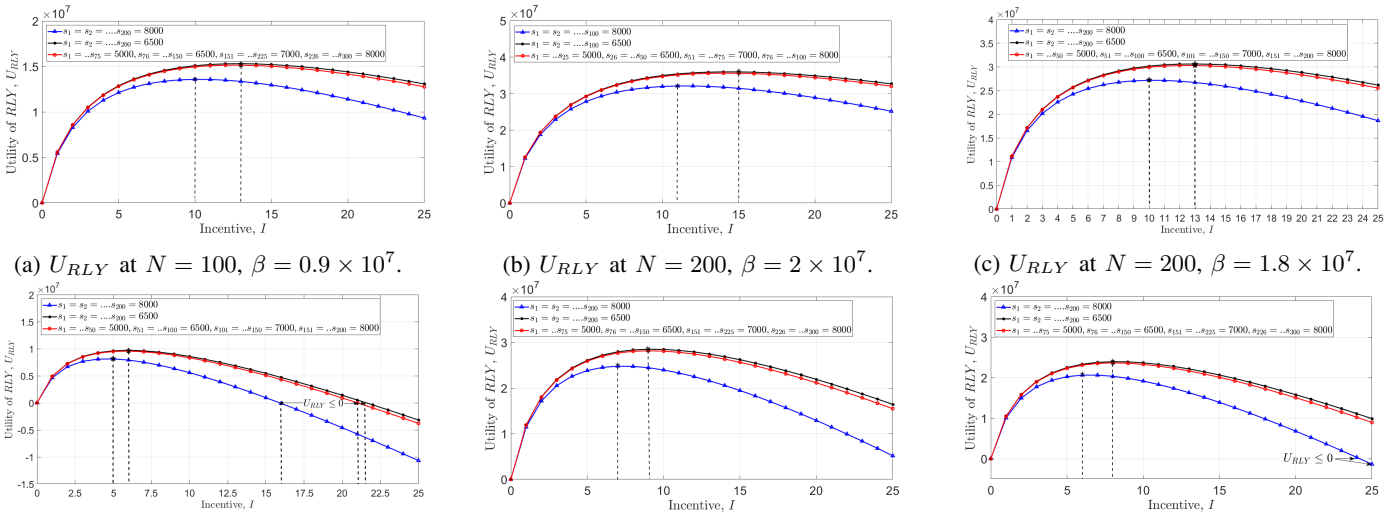
Fig. 13 shows the average time delay per hop in completing PoFL, PoQF [20] and PoS (Proof-of-Stake) [49] consensus in presence of both low and high vehicle density in the network, i.e., 100 vehicles and 300 vehicles . PoS is simulated such that it selects relay node on the basis of reputation of vehicle. A random reputation value following uniform distribution, ranging from 0 to 100 is assigned to each vehicle. The average time delay per hop of PoQF is proportional to number of vehicles and percentage of malicious vehicles in the network. This is because PoQF waits for a threshold number of votes to determine a relay node and the optimum threshold increases proportionally with number of vehicles and malicious vehicles percentage. Time delay of PoS changes according to number of vehicles due to more time required in accessing large amount of reputation values but it is independent of percentage of malicious vehicles. PoFL is run by each vehicle simultaneously and therefore its time delay is independent of both number of vehicles and percentage of malicious vehicles. On an average, PoFL is 65.2% faster than PoQF in relay selection and is more suitable for time-critical emergency situations. As a trade-off, PoQF only involves Quality Factor calculations but PoFL is based on a computationally expensive FL process with multiple iterations. Compared to PoS, PoFL is 15.74% faster when there are 300 vehicles but 18.93% slower when there are 100 vehicles. This is because PoS consumes time only in accessing the blockchain to find reputation of vehicles. The access time increases when there are more vehicles registered in a blockchain network. Although PoS with 100 vehicles outperforms both PoQF and PoFL, this faster consensus for block verification and addition cannot be run independently for appropriate relay selection, unlike PoQF and PoFL.

Table IX compares the asymptotic latency, communication and computation complexities of PoS, PoQF, PoFL and the blockchain-based FL process which is used to produce a global model for PoFL. Standard mathematical notations are used in Table IX, i.e., $\Omega(.)$, $O(.)$ and $\Theta(.)$ denote the order of *at least*, *at most* and *exactly* respectively. κ denotes consensus parameter and is unique to each algorithm, i.e., synchronization level in PoS, number of minimum votes required in PoQF and smallest dataset size in blockchain-based FL. The blockchain-based FL process has the highest computation complexity, as its computations depend on the size of dataset and number of vehicles submitting local models. Also, its communication complexity is proportional to the number of vehicles sharing their local models and its latency depends on the time of training a local model, which is proportional to the dataset size. However, PoFL, resulted from a blockchain-based FL process, outperforms PoS and PoQF because its latency, communication and computation complexities are independent of N and κ. It is therefore a highly scalable solution once the FL process is completed.



(a) $N_{RLY} = 6$. (b) $N_{RLY} = 8$.

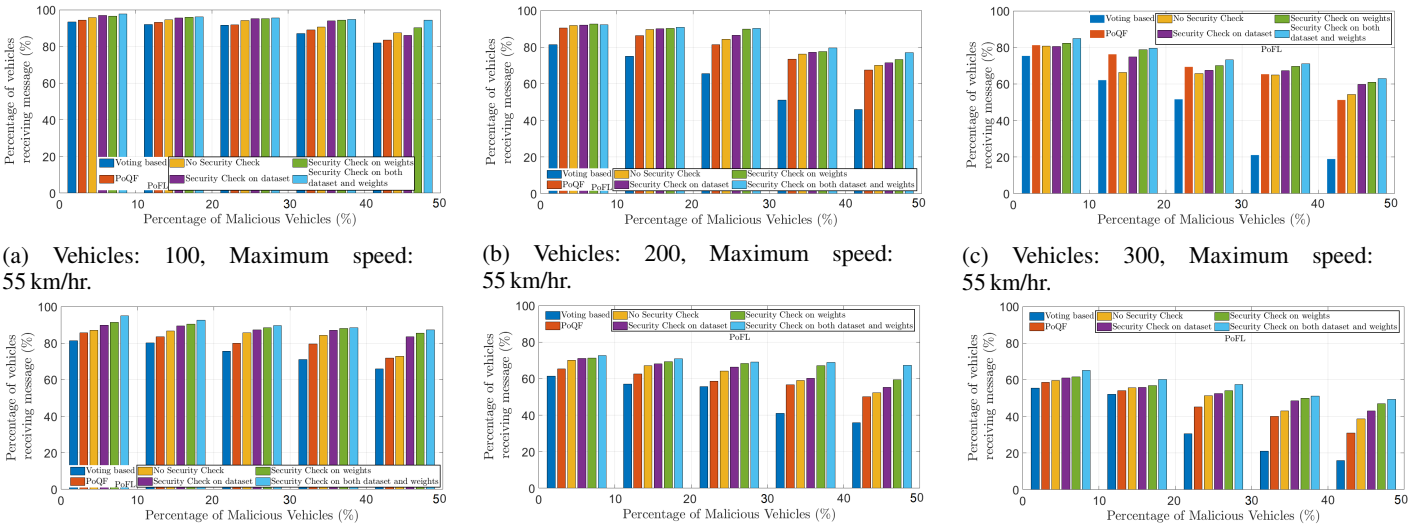
Fig. 10: U_i with equilibrium points (*) of Stackelberg Game.



(a) U_{RLY} at $N = 100$, $\beta = 0.9 \times 10^7$. (b) U_{RLY} at $N = 200$, $\beta = 2 \times 10^7$. (c) U_{RLY} at $N = 200$, $\beta = 1.8 \times 10^7$.

(d) U_{RLY} at $N = 200$, $\beta = 0.9 \times 10^7$. (e) U_{RLY} at $N = 300$, $\beta = 2 \times 10^7$. (f) U_{RLY} at $N = 300$, $\beta = 1.8 \times 10^7$.

Fig. 11: U_{RLY} with equilibrium points (*) of Stackelberg Game.



(a) Vehicles: 100, Maximum speed: 55 km/hr. (b) Vehicles: 200, Maximum speed: 55 km/hr. (c) Vehicles: 300, Maximum speed: 55 km/hr.

(d) Vehicles: 100, Maximum speed: 110 km/hr. (e) Vehicles: 200, Maximum speed: 110 km/hr. (f) Vehicles: 300, Maximum speed: 110 km/hr.

Fig. 12: Message delivery ratio.

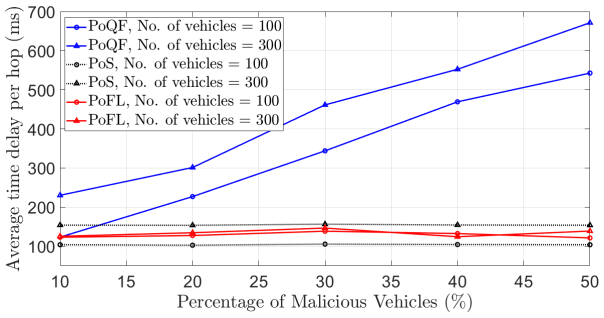


Fig. 13: Average time delay per hop.

TABLE IX: Comparison of Asymptotic Complexities.

Consensus	Latency	Communication	Computation
PoS	$\Omega(\kappa)$	$\Theta(1)$	$\Theta(1)$
PoQF	$\kappa O(1)$	$O(N)$	$\Theta(1)$
PoFL	$\Theta(1)$	$\Theta(1)$	$\Theta(1)$
Blockchain-based FL	$\Omega(\kappa)$	$\Omega(N)$	$N\Omega(\kappa)$

VI. Conclusion

In this paper, we have proposed a decentralized FL based message dissemination, governed by blockchain. The theoretical and practical performance of uploading local models using blockchain is compared with a centralized approach without blockchain. The proposed FL with blockchain can be considered as a faster approach since it results in more local models uploaded within a given time as compared to a solution without blockchain. Smart contract based security checks are proposed to detect adversary, which result in lower MSE in less number of iterations achieved by global model than FL without security check, after 100 iterations. Compared with other blockchain approaches suitable for relay selection in vehicular networks, the proposed solution is highly scalable, 65.2% faster and at least 8.2% more efficient in message dissemination approach. It also preserves privacy of neighbour vehicles, unlike other relay selection approaches. An economic model for blockchain based FL is also proposed and analyzed using Stackelberg game to determine optimal data size and incentive which result in the best response strategy of vehicles. Message dissemination and relay selection can further be improved in future work by including cross-layer information in dataset, obtained from physical and MAC layers.

References

[1] S. Yaqoob, A. Ullah, M. Akbar, M. Imran, and M. Shoaib, "Congestion avoidance through fog computing in internet of vehicles," *J. Amb. Intel. Hum. Comp.,* vol. 10, no. 10, pp.3863-3877, Feb. 2019.

[2] Y. Shi, L. Lv, H. Yu, L. Yu and Z. Zhang, "A Center-Rule-Based Neighborhood Search Algorithm for Roadside Units Deployment in Emergency Scenarios," *Mathematics,* vol. 8, no. 10, p.1734, Oct. 2020.

[3] M.F. Feteiha and M.H. Ahmed, "Multihop best-relay selection for vehicular communication over highways traffic," *IEEE Trans. Veh. Technol.,* vol. 67, no. 10, pp.9845-9855, Oct. 2018.

[4] A. Hawbani, E. Torbosh, W. Xingfu, P. Sincak, L. Zhao and A. Y. Al-Dubai, "Fuzzy based distributed protocol for vehicle to vehicle communication," *IEEE Trans. Fuzzy Syst,* Dec. 2019.

[5] M.E. Morocho-Cayamcela, H. Lee and W. Lim, "Machine Learning to Improve Multi-hop Searching and Extended Wireless Reach-ability in V2X," *IEEE Commun. Lett.,* vol 24, no. 7, pp. 1477-1481, Jul. 2020.

[6] W.K. Lai, M.T. Lin and Y.H. Yang, "A machine learning system for routing decision-making in urban vehicular ad hoc networks," *Int. J. Distrib. Sens. N.,* vol. 11, no. 3, p.374391, Mar. 2015.

[7] A. Mchergui, T. Moulahi and S. Nasri, "Relay Selection Based on Deep Learning for Broadcasting in VANET," in *Proc. 15th Int. Wireless Commun. Mobile Comput. Conf.*, Tangier, Morocco, Jun. 2019, pp. 865-870.

[8] W. He, G. Yan, and L. D. Xu, "Developing vehicular data cloud services in the IoT environment," *IEEE Trans. Ind. Informat.,* vol. 10, no. 2, pp. 1587–1595, May 2014.

[9] H. B. McMahan, E. Moore, D. Ramage, S. Hampson, and B. A. Y. Arcas, "Communication-efficient learning of deep networks from decentralized data," in *Proc. 20th Int. Conf. on Artificial Intelligence and Statistics,* Fort Lauderdale, FL, USA, Apr. 2017, pp. 1–11.

[10] W.Y.B. Lim, N.C. Luong, D.T. Hoang, Y. Jiao, C. Liang, Q. Yang, D. Niyato and C. Miao, "Federated learning in mobile edge networks: A comprehensive survey," *IEEE Commun. Surveys Tuts.,* vol. 22, no. 3, pp. 2031-2063, Apr. 2020.

[11] J. Kang, Z. Xiong, D. Niyato, S. Xie and J. Zhang, "Incentive mechanism for reliable federated learning: A joint optimization approach to combining reputation and contract theory," *IEEE Internet Things J.,* vol. 6, no. 6, pp.10700-10714, Dec. 2019.

[12] Y. Lu, X. Huang, K. Zhang, S. Maharjan and Y. Zhang, "Low-Latency Federated Learning and Blockchain for Edge Association in Digital Twin Empowered 6G Networks," *IEEE Trans. Industr. Inform.*, vol. 17, no. 7, pp. 5098-5107, July 2021.

[13] S. Nakamoto. (2008). *Bitcoin: A Peer-to-Peer Electronic Cash System.* [Online]. Available: https://bitcoin.org/bitcoin.pdf

[14] S.R. Pokhrel and J. Choi, "Federated learning with blockchain for autonomous vehicles: Analysis and design challenges," *IEEE Trans. Comm,* vol. 68, no. 8, pp. 4734-4746, Aug. 2020.

[15] K. Christidis and M. Devetsikiotis, "Blockchains and smart contracts for the Internet of Things," *IEEE Access,* vol. 4, pp. 2292–2303, May 2016.

[16] Y. Lu, X. Huang, K. Zhang, S. Maharjan and Y. Zhang, "Communication-Efficient Federated Learning and Permissioned Blockchain for Digital Twin Edge Networks," *IEEE Internet Things J.,* vol. 8, no. 4, pp. 2276-2288, 15 Feb.15, 2021.

[17] D. Mingxiao, M. Xiaofeng, Z. Zhe, W. Xiangwei and C. Qijun, "A review on consensus algorithm of blockchain," *Proc. IEEE Int. Conf. Systems, Man, and Cybernetics,* Banff, AB, Dec. 2017, pp. 2567-2572.

[18] V. Bagaria, S. Kannan, D. Tse, G. Fanti and P. Viswanath, "Prism: Deconstructing the Blockchain to Approach Physical Limits," in *Proc. ACM SIGSAC Conference on Computer and Communications Security,* New York, USA, Nov. 2019, pp. 585-602.

[19] W. Dong, Y. Li, R. Hou, X. Lv, H. Li and B. Sun, "A Blockchain-Based Hierarchical Reputation Management Scheme in Vehicular Network," in *Proc. IEEE Global Commun. Conf.*, Waikoloa, HI, USA, Dec. 2019, pp. 1-6.

[20] F. Ayaz, Z. Sheng, D. Tian and Y. L. Guan, "A Proof-of-Quality-Factor (PoQF) based Blockchain and Edge Computing for Vehicular Message Dissemination," *IEEE Internet Things J.*,vol. 8, no. 4, pp. 2468-2482, Feb. 2021.

[21] D. Cao, B. Zheng, B. Ji, Z. Lei and C. Feng, "A robust distance-based relay selection for message dissemination in vehicular network," *Wireless Networks,* vol. 26 no. 3, pp.1755-1771, Oct. 2018.

[22] S.A. Rashid, L. Audah, M. M. Hamdi and S. Alani, "Prediction Based Efficient Multi-hop Clustering Approach with Adaptive Relay Node Selection for VANET," *J. Commun.,* vol. 15, no. 4, pp. 332-334, Apr. 2020.

[23] F. Ayaz, Z. Sheng, D. Tian, Y. L. Guan and V. Leung, "A voting blockchain based message dissemination in vehicular ad-hoc networks (VANETs)," in *Proc. IEEE Int. Conf. Commun.*, Dublin, Ireland, Jun. 2020, pp. 1-6.

[24] F. Ayaz, Z. Sheng, D. Tian and V. Leung, "Blockchain-Enabled Security and Privacy for Internet-of-Vehicles," in *Internet of Vehicles and its Applications in Autonomous Driving,* N. Gupta, A. Prakash, R. Tripathi, Cham, Switzerland: Springer, Sep. 2020, pp. 123-148.

[25] C. Wu, S. Ohzahata, Y. Ji and T. Kato, "Joint fuzzy relays and network-coding-based forwarding for multihop broadcasting in VANETs," *IEEE Trans. Intell. Transp. Syst.,* vol. 16, no. 3, pp.1415-1427, Jun. 2015.

[26] K. Yang, Y. Shi, Y. Zhou, Z. Yang, L. Fu and W. Chen, "Federated Machine Learning for Intelligent IoT via Reconfigurable Intelligent Surface," *IEEE Netw.,* vol. 34, no. 5, pp. 16-22, Oct. 2020.

[27] K. Xiong, S. Leng, C. Huang, C. Yuen and Y. L. Guan, "Intelligent Task Offloading for Heterogeneous V2X Communications," *IEEE Trans. Intell. Transp. Syst.,* Aug. 2020.

[28] S. Samarakoon, M. Bennis, W. Saad, and M. Debbah, "Distributed federated learning for ultra-reliable low-latency vehicular communications," *IEEE Trans. Comm,* vol. 68, no. 2, pp. 1146–1159, Nov. 2019.

[29] M. Rihan, M. Elwekeil, Y. Yang, L. Huang, C. Xu and M.M. Selim, "Deep-VFog: When Artificial Intelligence Meets Fog Computing in V2X," *IEEE Syst. J.,* Aug. 2020.

[30] Y.M. Saputra, D.T. Hoang, D.N. Nguyen, E. Dutkiewicz, M.D. Mueck and S. Srikanteswara, "Energy demand prediction with federated learning for electric vehicle networks," in *Proc. IEEE Global Comm. Conf.,* Waikoloa, HI,USA, Dec. 2019, pp. 1-6.

[31] S. R. Pokhrel and J. Choi, "Federated learning with blockchain for autonomous vehicles: Analysis and design challenges," *IEEE Trans. Comm.,* vol. 68, no. 8, pp. 4734-4746, Aug. 2020.

[32] K. Tan, D. Bremner, J.L. Kernec and M. Imran, "Federated Machine Learning in Vehicular Networks: A summary of Recent Applications," in *Proc. IEEE Int. Conf. on UK-China Emerging Technol.,* Glasgow, UK, Aug. 2020, pp. 1-4.

[33] D. Ye, R. Yu, M. Pan and Z. Han, "Federated learning in vehicular edge computing: A selective model aggregation approach," *IEEE Access,* vol. 8, pp.23920-23935, Jan. 2020.

[34] Y. Sarikaya and O. Ercetin, "Motivating workers in federated learning: A Stackelberg game perspective," *IEEE Netw. Lett.,* vol. 2, no. 1, pp.23-27, Oct. 2019.

[35] S.R. Pandey, N.H. Tran, M. Bennis, Y.K. Tun, A. Manzoor and C.S. Hong, "A crowdsourcing framework for on-device federated learning," *IEEE Trans. on Wirel. Commun.,* vol. 19, no. 5, pp.3241-3256, Feb. 2020.

[36] R. Stanica, E. Chaput and A.-L. Beylot, "Local density estimation for contention window adaptation in vehicular networks," in *Proc. 22nd IEEE Int. Symp. Pers. Indoor Mobile Radio Commun.*, Toronto, ON, Canada, Sep. 2011, pp. 730–734.

[37] F.T. Liu, K.M. Ting and Z.H. Zhou, "Isolation forest," in *Proc. 8th IEEE Int. Conf. on Data Mining,*, Pisa, Italy, Dec. 2008, pp. 413-422.

[38] G. A. Susto, A. Beghi and S. McLoone, "Anomaly Detection through on-line Isolation Forest: An Application to Plasma Etching," in *Proc. 28th Annual SEMI Advanced Semiconductor Manufacturing Conf.*, Saratoga Springs, NY, USA, May 2017, pp. 89–94.

[39] K. Emara, "Safety-Aware Location Privacy in VANET: Evaluation and Comparison," *IEEE Trans. Veh. Technol.*, vol. 66, no. 12, pp. 10718-10731, Dec. 2017.

[40] J. Zhou et al., "Security-Critical Energy-Aware Task Scheduling for Heterogeneous Real-Time MPSoCs in IoT," *IEEE Trans. Services Comput.*, vol. 13, no. 4, pp. 745-758, Aug. 2020.

[41] A. W. Roscoe. (2021). *Temporal Signature in the Blockchain.* [Online]. Available: https://blockchain.univ.ox.ac.uk/wp-content/uploads/2021/05/Bill-Roscoe-Temporal-Signature.pdf

[42] K. Wei et al., "Federated Learning With Differential Privacy: Algorithms and Performance Analysis," *IEEE Trans. Inf. Forensics Secur.,* vol. 15, pp. 3454-3469, Apr. 2020.

[43] Y. Sun, L. Zhang, G. Feng, B. Yang, B. Cao, and M. A. Imran, "Blockchain-Enabled Wireless Internet of Things: Performance Analysis and Optimal Communication Node Deployment," *IEEE Internet Things J.*, vol. 6, no. 3, pp. 5791-5802, Jun. 2019.

[44] S. Kim, "Impacts of Mobility on Performance of Blockchain in VANET," *IEEE Access,* vol. 7, pp. 68646-68655, Jun. 2019.

[45] Z. Hou, H. Chen, Y. Li and B. Vucetic, "Incentive Mechanism Design for Wireless Energy Harvesting-Based Internet of Things," *IEEE Internet Things J.*, vol. 5, no. 4, pp. 2620-2632, Aug. 2018.

[46] F. Li, H. Yao, J. Du, C. Jiang and Y. Qian, "Stackelberg Game-Based Computation Offloading in Social and Cognitive Industrial Internet of Things," *IEEE Trans. Industr. Inform.,* vol. 16, no. 8, pp. 5444-5455, Aug. 2020.

[47] C. Sommer, R. German and F. Dressler, "Bidirectionally Coupled Network and Road Traffic Simulation for Improved IVC Analysis," *IEEE Trans. Mobile Comput.,* vol. 10, no. 1, pp. 3-15, Jan. 2011.

[48] T. E. Oliphant, "Python for Scientific Computing," *Comput. Sci. Eng.*, vol. 9, no. 3, pp. 10-20, May-June 2007.

[49] J. Kang, Z. Xiong, D. Niyato, D. Ye, D. I. Kim, and J. Zhao, "Toward Secure Blockchain-Enabled Internet of Vehicles: Optimizing Consensus Management Using Reputation and Contract Theory," *IEEE Trans. Veh. Technol.*, vol. 68, no. 3, pp. 2906–2920, Mar. 2019.

Ferheen Ayaz (Graduate Student Member, IEEE) received her B.E. and M.E. degree from NED University of Engineering and Technology, Pakistan, in 2010 and 2014, respectively. She is currently completing her Ph.D. degree with University of Sussex, UK. She is working as a Research Associate in University of Glasgow, UK. Her current research interests include security and privacy, vehicular communications, blockchain and machine learning.

Zhengguo Sheng (Senior Member, IEEE) received the B.Sc. degree from the University of Electronic Science and Technology of China, Chengdu, China, in 2006, and the M.S. and Ph.D. degrees from Imperial College London, London, U.K., in 2007 and 2011, respectively. He is currently a Reader with the University of Sussex, Brighton, U.K. Previously, he was with UBC, Vancouver, BC, Canada, as a Research Associate and with Orange Labs as a Senior Researcher. He has more than 120 publications. His research interests cover IoT, vehicular communications, and cloud/edge computing.

Daxin Tian (Senior Member, IEEE) is a professor in the School of Transportation Science and Engineering, Beihang University, Beijing, China. His current research interests include mobile computing, intelligent transportation systems, vehicular ad hoc networks, and swarm intelligence.

Yong Liang Guan (Senior Member, IEEE) obtained his PhD from the Imperial College London, UK, and Bachelor of Engineering with first class honors from the National University of Singapore. He is a Professor of Communication Engineering at the School of Electrical and Electronic Engineering, Nanyang Technological University, Singapore, where he now leads two industry collaboration labs (Continental-NTU Corporate Research Lab, and Schaeffler Hub for Advanced Research at NTU) and led the successful deployment of the campus-wide NTU-NXP V2X Test Bed. His research interests broadly include coding and signal processing for communication systems and data storage systems. He is an Editor for the IEEE Transactions on Vehicular Technology.